

Improving Remote Sensed Data Products Using Bayesian Methods for the Analysis of Computer Model Output

Robin Morris

USRA-RIACS

Athanasis Kottas and Matt Taddy

Department of Applied Mathematics and Statistics, University of California, Santa Cruz

Roberto Furfarò and Barry Ganapol

Aerospace and Mechanical Engineering Department, University of Arizona, Tucson

AISRP PI Meeting, May 2008

Why uncertainty analysis is important

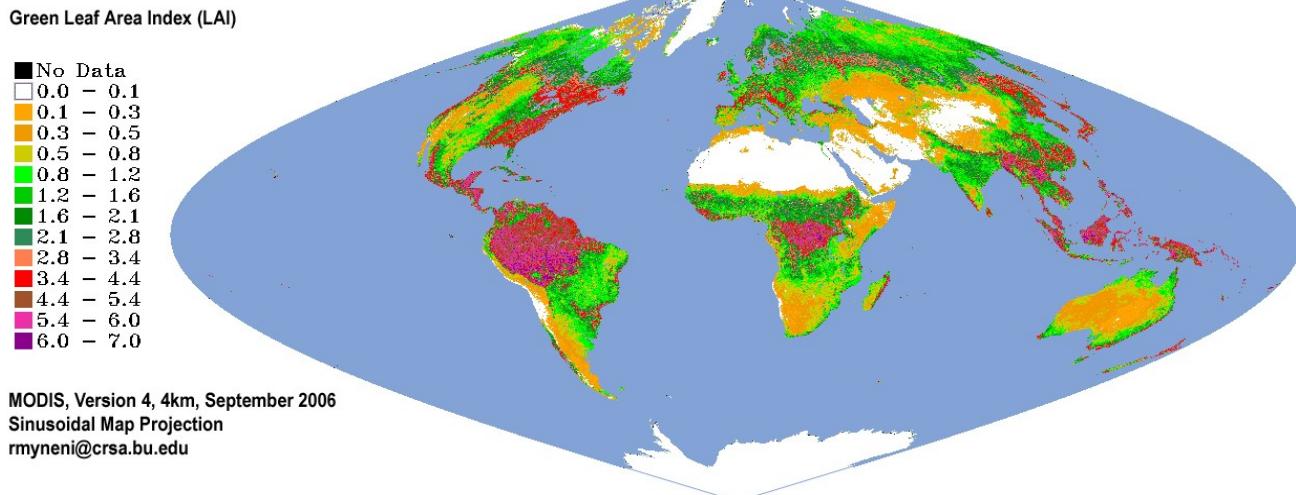
- Biosphere properties - estimation is extremely important
 - Critical for our understanding of the Earth's coupled systems
 - Estimation requires remote sensing (space-based) observations
 - NASA is committed to "*Us[ing] space-based observations to improve understanding and prediction of Earth system variability and change for climate, weather and natural hazards*".
 - Observation are integrated in large-scale models
 - How does Earth respond to changes in climate and chemical composition?
 - Example: CASA model uses LAI and FPAR as input parameters
-

Why uncertainty analysis is important (cont)

- **Analyzing and quantifying uncertainties is vital**
 - An estimate without its associated uncertainty has very diminished value
 - Uncertainties are important for both science and public policy
 - Special issue on **Global Land Product Validation** (Morisette et. al. 2006):
 - “*...Users need access to quantitative information on product uncertainties...*”
 - “*...Making quantified accuracy information available to the users can ultimately provide developers the necessary feedback for improving the products...*”
 - Estimation of vegetation properties done via **Radiative Transfer Models (RTM)**
 - Complex inference and prone to inversion problems

Example: MODIS/LAI algorithm

- **MOD-15 Algorithm** for retrieval of Leaf Area Index (LAI)
 - RTM inversion based on **Look-up table approach**
 - *Back-up algorithm uses an empirical NDVI-reflectance relationship*
 - **Algorithm has improved** continuously over the years
 - *Use of a better biome map: reducing the uncertainty in that input*
 - *Better atmospheric correction*
 - *Improved models for surface reflectance as function of the biome*
 - More work to be done: The improvements have reduced the uncertainty in results but **have not improved uncertainty quantification**
 - *Statistical identification of the uncertainty sources is required*



Statistical Sensitivity Analysis of RTMs

- Analyze the effects of the inputs to an RTM in terms of sensitivity of RTM's output to each input
 - Global sensitivity analysis rather than error analysis around a fixed point
 - Employ a Leaf-Canopy radiative transfer model (LCM) as surrogate for the RTM used by the MODIS algorithm
 - Main Effects
 - Graphically show the relative importance of each input
 - Sensitivity Indices
 - Measure of the expected amount by which each uncertainty in the output would be reduced if the true value were known

Statistical Sensitivity Analysis of RTMs (cont)

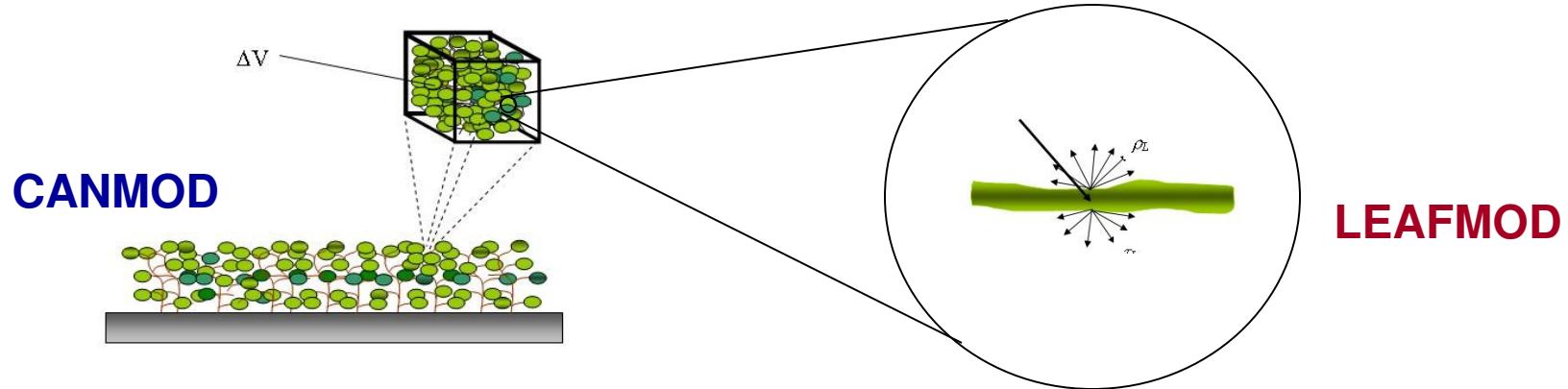
- Computing the main effects and the sensitivity indices requires evaluating multidimensional integrals
 - Using the LCM directly is computationally prohibitive
 -
- Approximate the LCM by a Gaussian Process (GP)
 - GP model is a very flexible non parametric approach to function approximation
 - *GP models can be constructed with a few, carefully chosen runs of the LCM*
 - *The integrals for the main effects and sensitivity indices become analytically tractable*

LCM: modelling the radiative regime in plant canopies

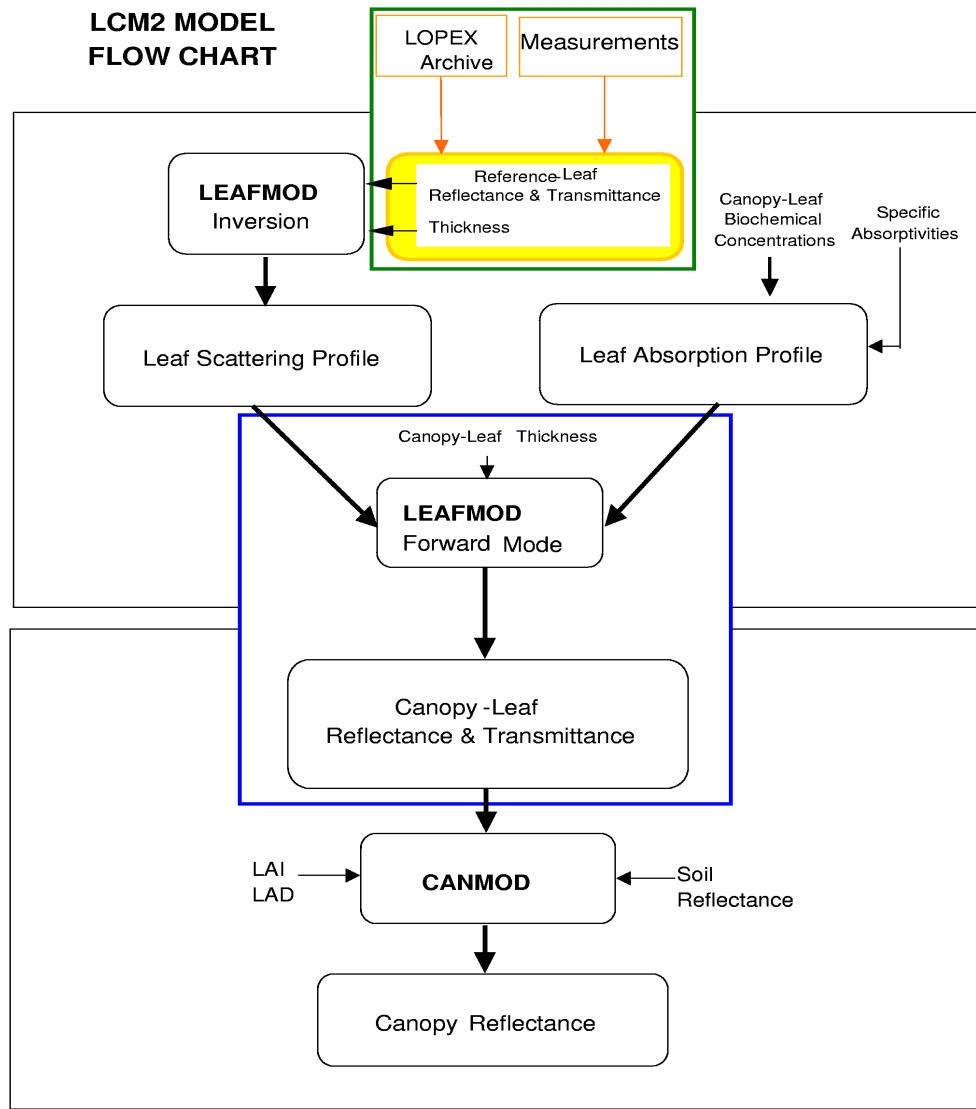
Coupled Leaf-Canopy Model (LCM): a true **radiative transport model** to simulate the radiative field in vegetation canopies

Approach used in designing LCM:

- First principles direct application: **Balance of photons**
- **Two level of description:** Leaf and Canopy model
- **Biochemistry** included as key element affecting vegetation optics
- Connection of the models via **leaf optical properties**



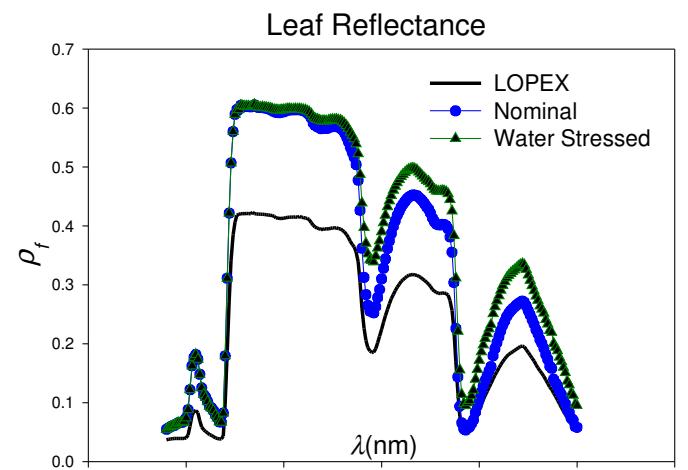
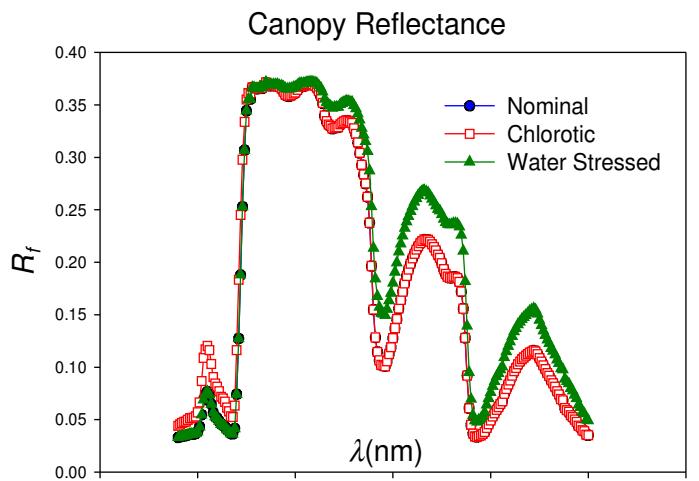
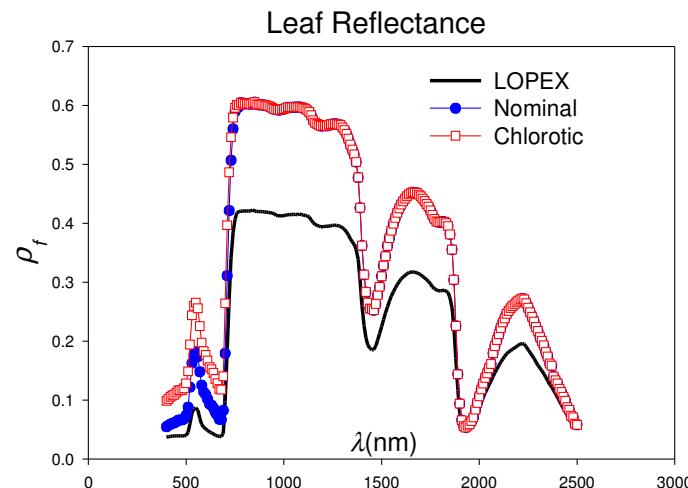
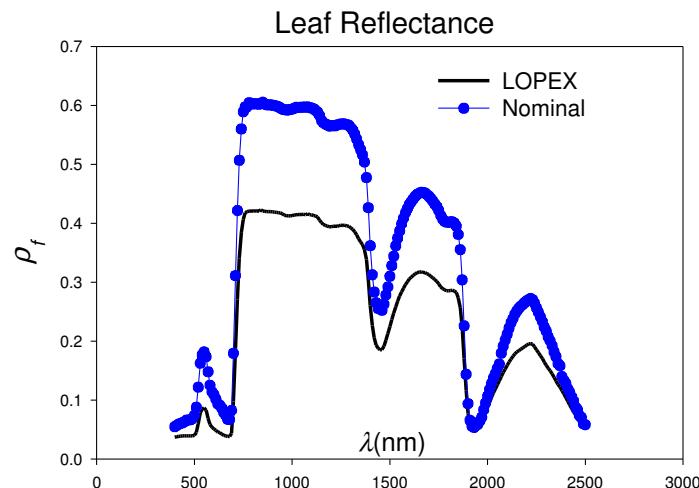
Leaf-Canopy Model: LCM2 flowchart



Model Inputs

- LAI
- Chlorophyll
- water fraction
- protein
- lignin/cellulose
- thickness
- soil

Leaf-Canopy Model (LCM): models the effects of leaf chemistry



Sensitivity Analysis: Main Effects and Sensitivity Indices

- **Goal:** Determine how the variation in the LCM output can be apportioned amongst inputs
 - Global sensitivity analysis is of interest (vs. local analysis)
 - *How the output changes as all inputs vary continuously*
 - Probability distribution ($H(v)$) for the inputs is required
 - *$H(v)$ is problem dependent and encodes correlation between variables*
 - *Uniform distribution used for parameters varying in a specific range*
- Decompose the response of the model to inputs in expected value, main effects and multiple interaction terms

$$\begin{aligned}y = f(v) &= E(Y) + \sum_{i=1}^d z_i(v_i) + \sum_{i < j} z_{i,j}(v_i, v_j) + \dots \\&\quad + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d)\end{aligned}$$

Sensitivity Analysis: Main Effects and Sensitivity Indices (cont)

- **Main Effects**
 - Give information about the **combined influence** of two or more inputs taken together
 - Plotting mean effects gives a **visual impression** of their relative importance
- **Sensitivity indices**
 - Expected amounts by which the uncertainty in the output is reduced if **the i-th true value were known** (Normalized)
 - The sum of first and higher order indices is **unity**

Main effects:

$$\begin{aligned}z_i(v_i) &= E(Y|v_i) - E(Y) \\&= \int_{v_{-i}} f(v)dH(v_{-i}|v_i) - E(Y)\end{aligned}$$

Variance and sensitivities:

$$\begin{aligned}V_i &= \text{Var}\{E(Y|v_i)\}. \\S_i &= V_i/\text{Var}(Y)\end{aligned}$$

Modeling LCM via Gaussian Processes

- **Gaussian Process Models:** Approximate a function by placing a distribution directly over the function space
 - The joint distribution of any finite set of points is a **multivariate Gaussian** hence analytically tractable
 - $E(f(v)) = \mu,$ $\text{Corr}(f(v), f(v'); \theta) = \exp\left(-\sum_{\ell=1}^d \frac{(v_\ell - v'_\ell)^2}{\gamma_\ell}\right)$
 - $\text{Var}(f(v)) = \sigma^2$ $\theta = (\gamma_1, \dots, \gamma_d),$
- Use the value as above to define the **joint probability distribution:**
$$p(y|\theta, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}|C|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu\mathbf{1}_n)^T C^{-1}(y - \mu\mathbf{1}_n)\right)$$
- **Training samples** are used to estimate the GP parameters using **maximum likelihood** and the log of the joint distribution:

$$\mathcal{L} = -\frac{1}{2\sigma^2}(y - \mu\mathbf{1}_n)^T C^{-1}(y - \mu\mathbf{1}_n) - \frac{1}{2} \log |C| - \frac{n}{2} \log(2\pi\sigma^2)$$

Modeling LCM via Gaussian Processes (cont)

- Once the **GP model parameters** are estimated, we compute the predictive **distribution sets** of new input conditioned by the training examples

- Predictive distribution are **Gaussian** for any new input v

- Mean:**

$$m \equiv m(v; \hat{\mu}, \hat{\theta}, d) = \hat{\mu} + r^T(v)C^{-1}(y - \hat{\mu}\mathbf{1}_n)$$

- Variance:**

$$S \equiv S(v; \hat{\mu}, \hat{\sigma}^2, \hat{\theta}, d) = \hat{\sigma}^2 (1 - r^T(v)C^{-1}r(v))$$

- Joint predictive distribution to two inputs (i.e. $(f(v), f(v'))$)**

$$w = \hat{\mu}\mathbf{1}_2 + R^T(v, v')C^{-1}(y - \hat{\mu}\mathbf{1}_n)$$

$$W = \hat{\sigma}^2 (B(v, v') - R^T(v, v')C^{-1}R(v, v'))$$

Using Gaussian Processes to compute main effects and sensitivity indices

- To compute **main effects** requires evaluating $E(Y | v_j)$ and $E(Y)$
 - Because GP predicts distributions we need expected values:

$$E^* \{E(Y | v_j)\} \quad E^* \{E(Y)\},$$

- For the **global mean**

$$E(Y) = \int_v f(v) \prod_{\ell=1}^d dH_\ell(v_\ell) \quad H(v) = \prod_{\ell=1}^d H_\ell(v_\ell)$$

$$E^* \{E(Y)\} = \int_v m(v) \prod_{\ell=1}^d dH_\ell(v_\ell) = \hat{\mu} + T^T C^{-1}(y - \hat{\mu} \mathbf{1}_n)$$

- For the **conditional mean**

$$E(Y | u_j) = \int_{\{v_\ell : \ell \neq j\}} f(v_1, \dots, u_j, \dots, v_d) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell)$$

$$E^* \{E(Y | u_j)\} = \int E(Y | u_j) dN(f(v_1, \dots, u_j, \dots, v_d); m, S)$$

$$= \int_{\{v_\ell : \ell \neq j\}} m(v_1, \dots, u_j, \dots, v_d) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell)$$

$$= \hat{\mu} + T_j^T (u_j) C^{-1}(y - \hat{\mu} \mathbf{1}_n),$$

Using Gaussian Processes to compute main effects and sensitivity indices (cont)

- Measure of **uncertainty** associated to the GP estimates

$$\text{Var}^* \{\mathbb{E}(Y | u_j)\} = \mathbb{E}^* \{(\mathbb{E}(Y | u_j))^2\} - (\mathbb{E}^* \{\mathbb{E}(Y | u_j)\})^2.$$

- After some **lengthy derivation** we get:

$$\text{Var}^* \{\mathbb{E}(Y | u_j)\} = \hat{\sigma}^2 (e - T_j^T(u_j) C^{-1} T_j(u_j))$$

- For the sensitivity indices the **expected variance** must be computed

- The **sensitivity indices** are defined as follow:

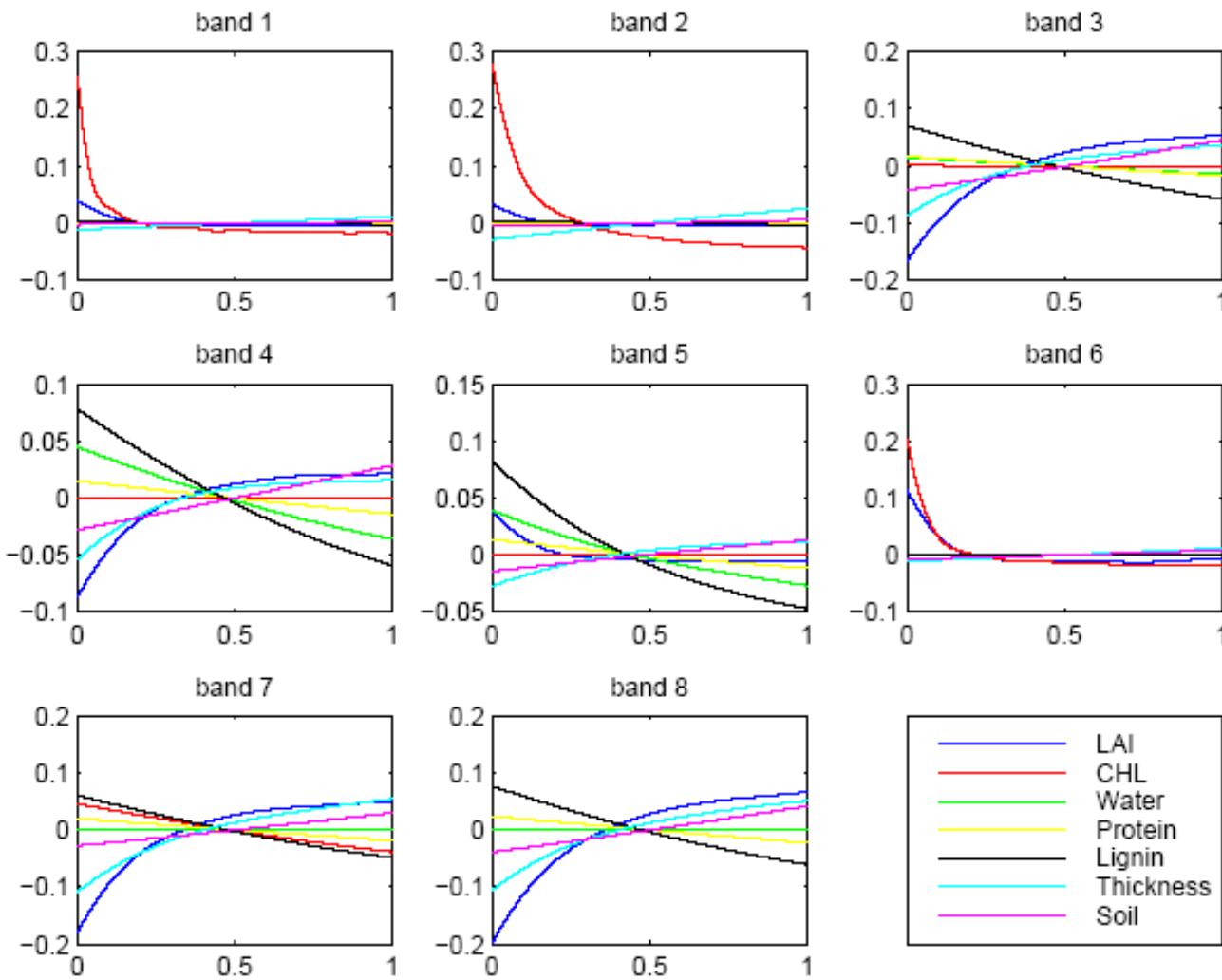
$$S_j = \frac{\text{Var}(\mathbb{E}(Y | u_j))}{\text{Var}(Y)}, \quad j = 1, \dots, d.$$

- A measure of their uncertainty is given by the ratio of:

$$\mathbb{E}^* \{\text{Var}(\mathbb{E}(Y | u_j))\} = \mathbb{E}^* \{ \mathbb{E} [(\mathbb{E}(Y | u_j))^2] \} - \mathbb{E}^* \{ (\mathbb{E}(Y))^2 \}$$

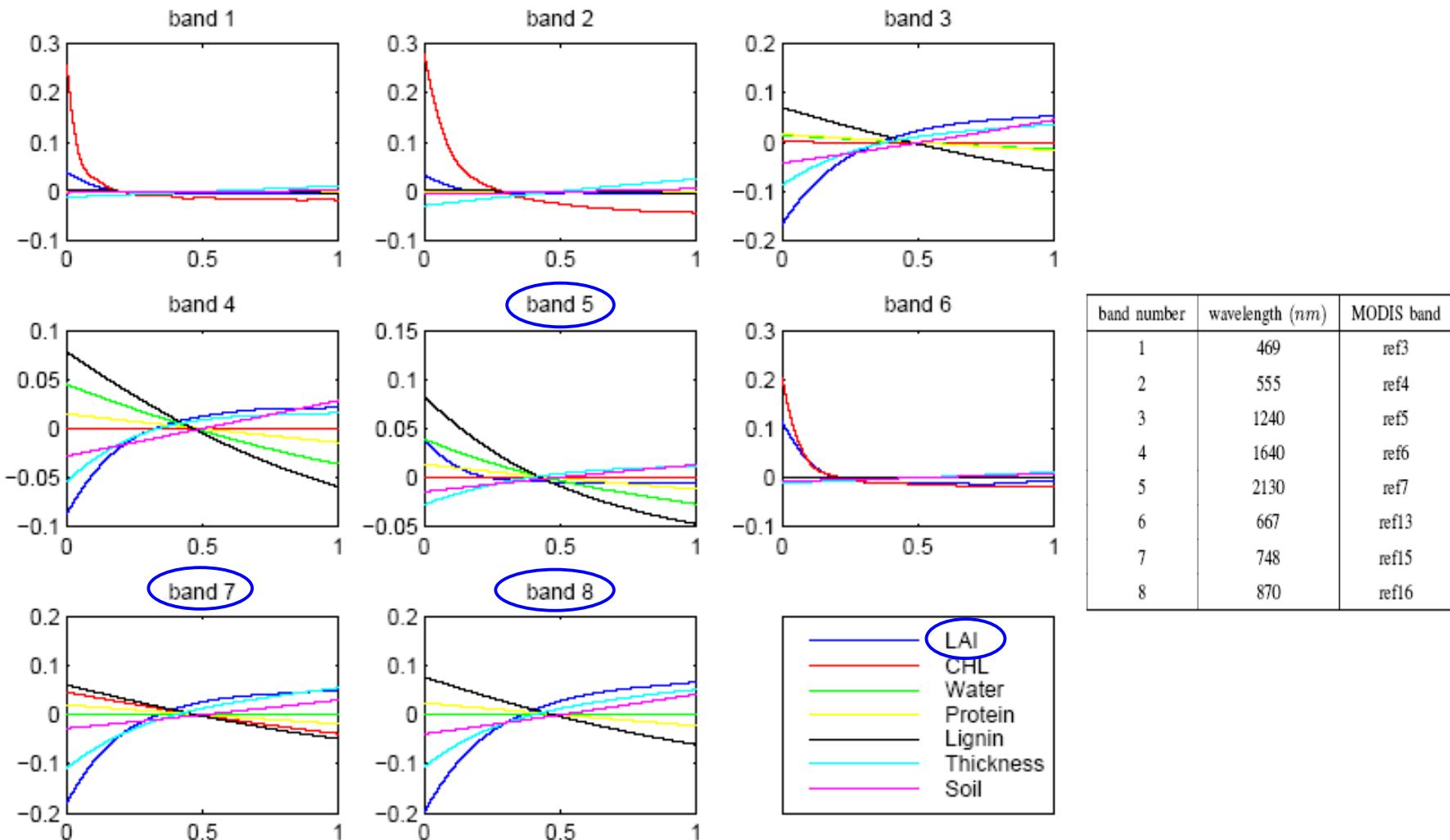
$$\mathbb{E}^* \{\text{Var}(Y)\} = \mathbb{E}^* \{ \mathbb{E}(Y^2) \} - \mathbb{E}^* \{ (\mathbb{E}(Y))^2 \}$$

Results: Main effects



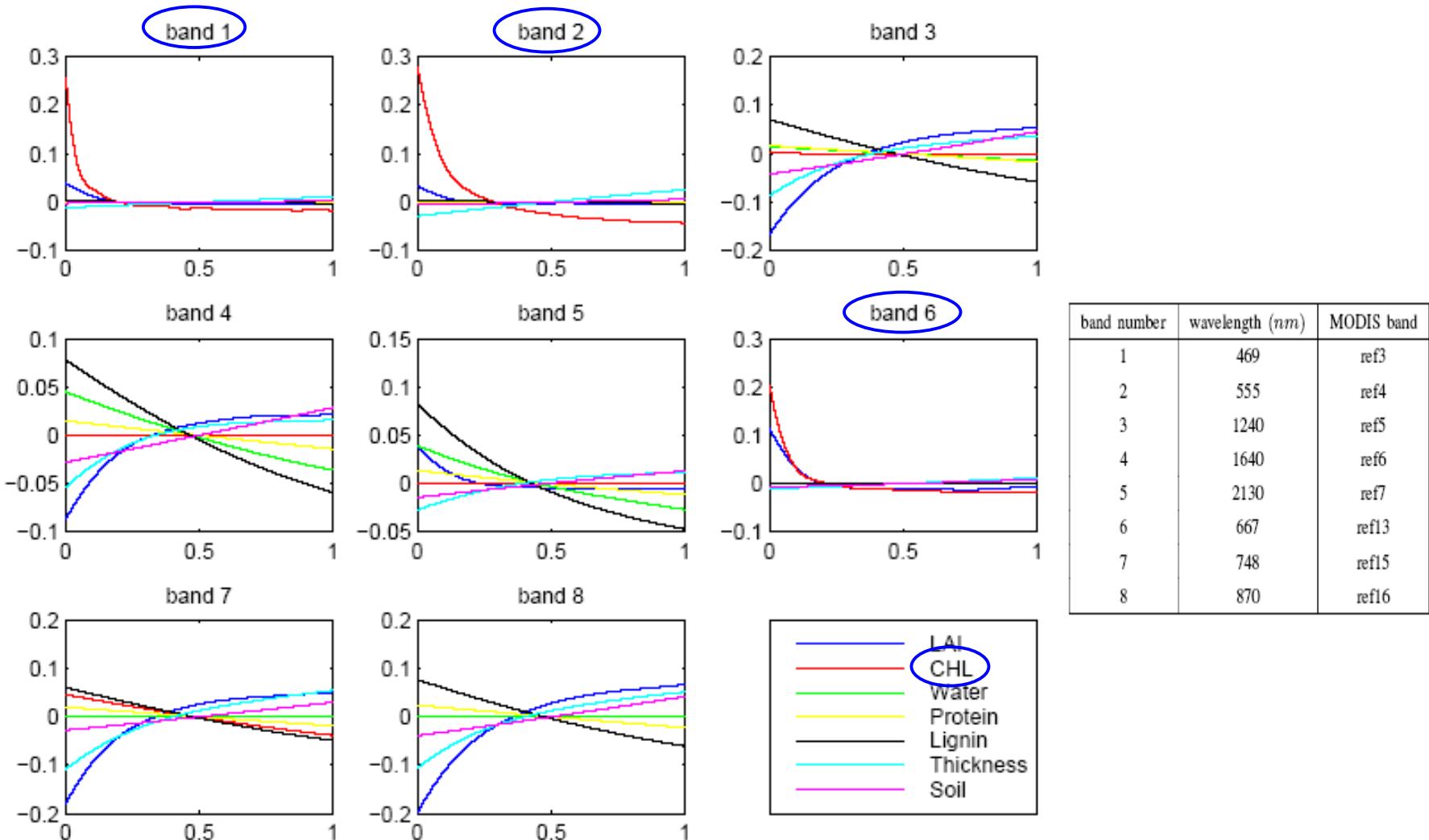
band number	wavelength (nm)	MODIS band
1	469	ref3
2	555	ref4
3	1240	ref5
4	1640	ref6
5	2130	ref7
6	667	ref13
7	748	ref15
8	870	ref16

Results: Main effects



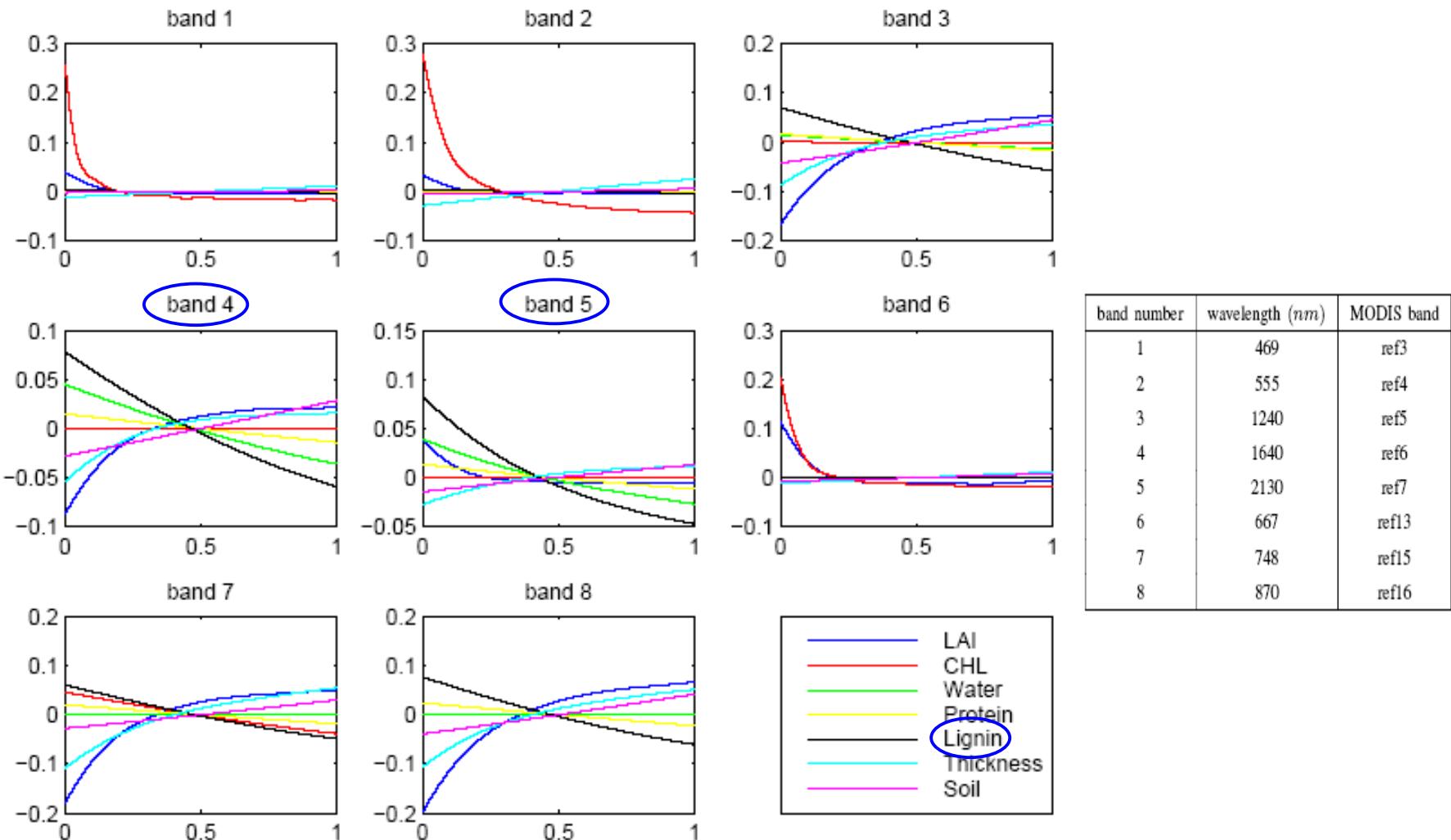
LAI: sensitive in NIR (bands 7,8); opposite effect in visible (band 5)

Results: Main effects



Chlorophyll: sensitive in visible (bands 1,2,6); disappears after “red-edge”

Results: Main effects



Lignin: sensitive in SWIR (bands 4,5); surprising to domain scientists

Results: Sensitivity indices

	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Results: Sensitivity indices

	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Results: Sensitivity indices

	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

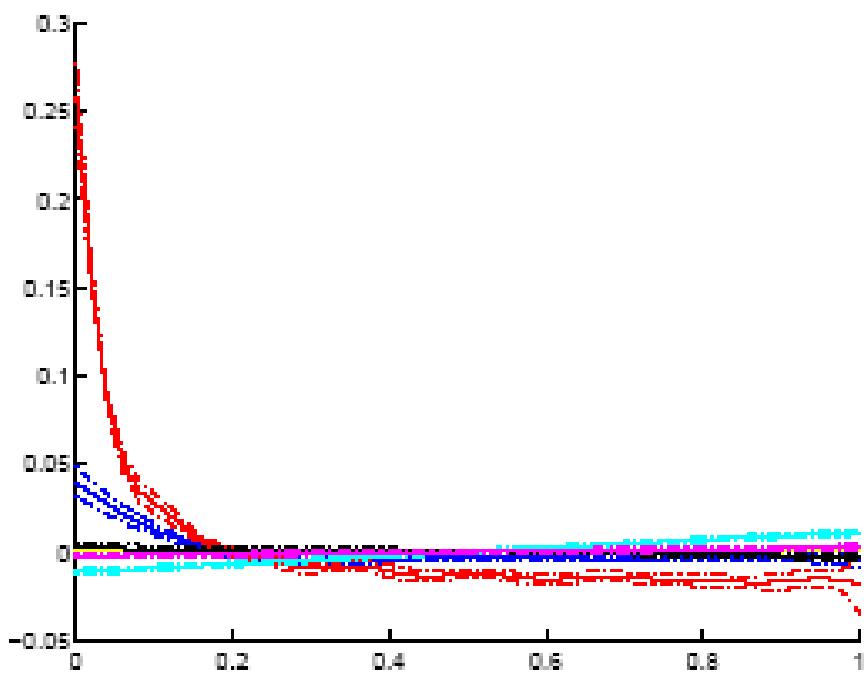
Results: Sensitivity indices

	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

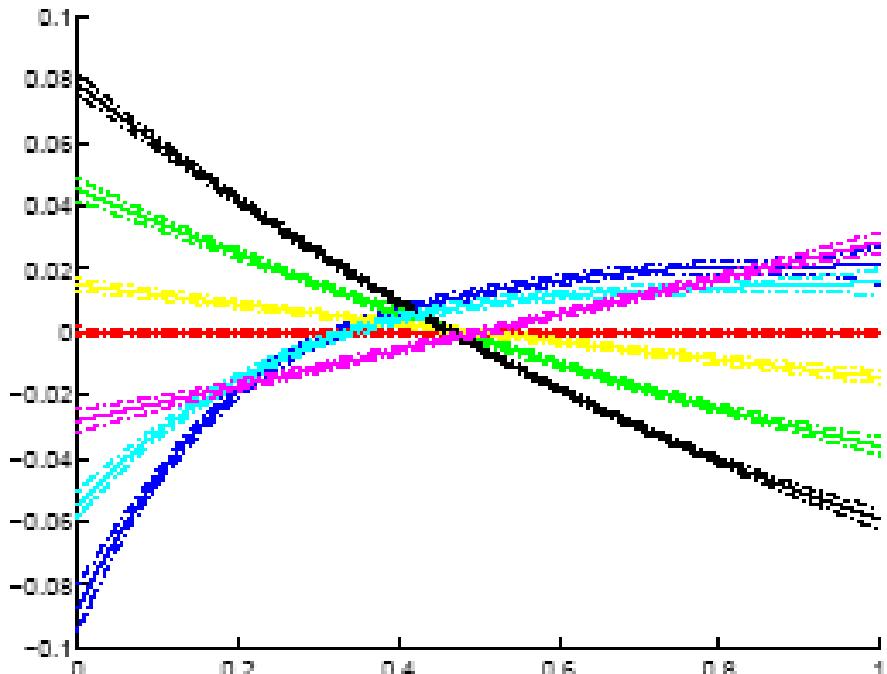
Results: Sensitivity indices

	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Results: Main effect and its uncertainties



Band 1



Band 4

Conclusions

- Results show that analyzing characteristics of RTMs used in remote sensing products is practical and important
 - *First attempt to put the uncertainty analysis in a Bayesian framework*
 - *Results confirmed similar sensitivity trend in LAI, Chlorophyll and soil brightness performed by Bacour et. al. 2002*
 - *Global sensitivity analysis is more important than local sensitivity*
- Uncertainty analysis gives useful information
 - Level of accuracy needed in model inputs
 - Which inputs can actually be observed remotely
 - Can guide data collection to reduce uncertainty
 - Can guide further development of the RTMs themselves
-

Highlights

- “A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models”, accepted for publication in IEEE Transactions on Geoscience and Remote Sensing
- “An Analysis of the Uncertainties in Radiative Transfer Models Used in Remote Sensed Data Product Generation”, NASA Science Technology Conference, 2007
- “Global Sensitivity Analysis of Leaf-Canopy Radiative Transfer Models for Analysis and Quantification of Uncertainties in Remote Sensed Data Product Generation”, presented at AGU Fall Meeting, 2007
- “A Gaussian Process Approach to Quantifying the Uncertainty of Vegetation Parameters from Remote Sensing Observations”, AGU Fall Meeting, 2006

Work-in-progress/Future Plans

- Refining the LCM to compute Bi-directional reflectances
- Computing a true uncertainty on the sensitivity indices
- Developing a hierarchical GP model to incorporate LAD in a unified framework
- Developing an improved non-parametric function estimator for model approximation
- Model inversion from remote sensed data
- Validation using BIGFOOT and other data